Disambiguating form and lexical frequency effects in MEG responses using homonyms

Dylan Alexander Simon¹, Gwyneth Lewis², Alec Marantz¹,²

¹Department of Psychology, ²Department of Linguistics
New York University
6 Washington Pl.
New York, NY 10003

Short title: Homonym lexical resolution

Correspondence should be addressed to Dylan A. Simon

Telephone: 212-998-7605

Email: dylex@nyu.edu
Abstract

We present an MEG study of homographic homonym recognition in reading, aiming to distinguish between two theories of lexical access: the “early access” theory, which entails that lexical access occurs at early (pre 200 ms) stages of processing, and the “late access” theory, which interprets this early activity as orthographic word-form identification rather than genuine lexical access. A correlational analysis method was employed to examine effects of the lexical frequencies of distinct words that share the same orthography (homographs) on brain activity. We find support for the “late access” view, in that lexical frequency did not affect processing until after 300 ms, while earlier activation was primarily modulated by orthographic form frequency.
Introduction

Understanding spoken or written language is the process of transforming sensory stimuli into internal mental representations of meaning. A key component of this transformation is lexical resolution: the connection of orthographic or phonological representations of words to their semantic referents. The precise time-course of linguistic processing can be studied using neuroimaging techniques with high temporal precision such as EEG or MEG. The finding of a neural response that correlates with some lexical property of the word is often used as evidence that lexical access has begun by the time of that response. One such putative lexical property is the frequency of the word in a language. Word frequency has been well established as a factor speeding response time in a variety of experimental tasks (Scarborough, Cortese, & Scarborough, 1977). Embick, Hackl, Schaeffer, Kelepir, and Marantz (2001) found reliable correlates to the behavioral latency effect in the latency of MEG M350 responses, while others have found frequency effects in EEG N280 and N400 potentials and earlier MEG responses (Assadollahi & Pulvermüller, 2003; Hauk & Pulvermüller, 2004; King & Kutas, 1995). However, to make the interpretation unambiguous, confounding factors that may give rise to correlations between lexical and non-lexical properties must be carefully considered.

Similar frequency effects have been found using non-word stimuli, indicating that these results may be due to form frequencies reflecting common sequences of letters. One fMRI study focusing on the visual word form area (VWFA) in which many of these effects have been localized showed no significant difference between frequent quadragrams and real words (Vinckier et al., 2007). Various manipulations have been done to try to identify semantically sensitive components of responses in order to locate lexical access in the course of word recognition and resolve this frequency effect confound. Hauk, Davis, Ford, Pulvermüller, and Marslen-Wilson (2006) regressed EEG responses on correlates of various lexical properties, including a measure of semantic coherence based on the morphological family of each word. They found early effects of form frequency, by around 100 ms, and slightly later effects of a measure correlated with semantic coherence, concluding that
lexical resolution happens between 100 and 170 ms. On the other hand, Pylkkänen, Stringfellow, and Marantz (2002) provide evidence that the lexical process cannot have completed before 350 ms by manipulating two different but correlated lexical variables that affect behavioral lexical decision time. They found that phonotactic probability was associated with shorter M350 response latencies, while no effects were found from phonological neighborhood density, which under early access theories should have an inhibitory effect on post-lexical processing. Solomyak and Marantz (2009a) developed a novel technique for comparing overall form frequency to lexical frequency in MEG responses using non-homophonic homographs (e.g., the verb or noun wind) to disambiguate effects related to orthographic forms from those dependent on phonological resolution. They found early effects of form frequency on a response in the inferior occipitotemporal region (associated with the MEG M170) around 150 ms, while effects from the phonologically distinct word frequencies were only found at a much later response in the superior temporal and Sylvian Fissure region after 300 ms (MEG M350).

We focus on a specific kind of linguistic ambiguity to resolve this dilemma: homographic homonymy, where distinct word meanings have identical orthographies, such as the verb and noun peer. This is distinguished from polysemy, in which a single meaning has multiple related senses. There is evidence that these two types of ambiguity have distinct effects on processing, both in response time (Rodd, Gaskell, & Marslen-Wilson, 2002), and in the MEG M350 response (Beretta, Fiorentino, & Poeppel, 2005). There is also some evidence that homonyms have separate neural representations, each of which may cover many polysemic senses (Pylkkänen, Llinás, & Murphy, 2006; Tamminen, Cleland, Quinlan, & Gaskell, 2006). In one class of network models, senses of a word provide excitatory activation to other senses of the same word, while distinct meanings have inhibitory connections between them just as distinct words do (Rodd, Gaskell, & Marslen-Wilson, 2004). Under these models, the latency of activation reflects the settling time of the network, while the magnitude reflects the extent of the activity, both of which will be increased by conflicting activations. Accordingly, when an isolated homograph is presented, all representations are initially activated,
but activation will not peak until a single meaning entry is selected.

Since homonyms have distinct meanings, each meaning can be said to have its own frequency in the language, and these lexical frequencies will sum to the frequency of the shared form. Additionally, there is a range of ambiguity among homonyms from balanced, where all meanings are similarly likely as in *pawn*, to unbalanced, where one meaning is dominant as in *down*. Under the network interpretation, more balance among the meanings of a given homonym results in more competition between meanings, and thus more prolonged and stronger network activity. This is consistent with the many findings of meaning ambiguity disadvantage (Klepousniotou & Baum, 2007) and similarly suggests increased neural activity in resolving more balanced words.

In this study, we aimed to exploit these balance differences to clearly disambiguate lexical and word form frequency effects. We tested the effect of varying homographs along the meaning balance dimension while controlling for form frequency and other confounding factors. We looked for effects of this lexical distinction on the M130 and M170 MEG responses to visual words in previously identified letter string (LS) and visual word form (VWF) areas to determine whether frequency effects shown in this time range (100–200 ms) reflect lexical or form variation. The definition of the time windows and regions of interest was motivated by previous work distinguishing two different early responses components to visual letter string stimuli: an earlier Type I response in occipital visual areas, and a left-lateralized Type II response in the inferior occipitotemporal region (Tarkiainen, Cornelissen, & Salmelin, 2002; Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999). These responses have been previously identified in MEG by the time window and magnetic field patterns and by field direction in both single-source dipole and distribution source cortical models (Solomyak & Marantz, 2009b). To quantify ambiguity, we used a measure of meaning entropy introduced by Twilley, Dixon, Taylor, and Clark (1994) which represents the information content of a meaning resolution. Since a measure of true relative lexical frequency was unavailable, the proportion of dictionary senses within each meaning was used as a correlate of this value. This same number of senses count was shown to have the correct correlations with frequency, as well as
a direct advantage for lexical access independent of frequency in earlier results (Rodd et al., 2002).

[Figure 1 about here.]

Methods

Participants

Thirteen right-handed native English speakers (eight female) with normal or corrected-to-normal vision participated in the experiment. The participants were at or above the undergraduate level and were presumed to be normal readers. All participants gave informed consent and received $15 for their participation. The study was approved by the New York University Committee on Activities Involving Human Subjects.

Materials

The stimuli consisted of 500 homographic English words (between 3 and 6 letters long) and an equal number of length-matched non-words. Word lists and properties were collated from the English Lexicon Project (Balota et al., 2002) and Wordsmyth Online Dictionary (Parks, Kennedy, & Broquist, 1998). For each word, we used the following measures from the English Lexicon Project: length in characters ($L$), summed bigram frequency ($B$), and log frequency in the HAL corpus ($F$). From the Wordsmyth Online Dictionary we extracted: the number of meanings ($M$) as the count of dictionary headword entries, and the number of senses within each meaning ($S_m$). All words were monomorphemic with no alternate spellings, phrasal dictionary entries, irregular capitalization, or acronyms. We also determined the morphological family (cohort) size and summed frequency across all entries in CELEX (Baayen, Piepenbrock, & Gulikers, 1995) containing the word as a root.
Given the number of meanings for each word, $M_n$, and the number of senses for each meaning, $S_{n,m}$, we used the sense counts to estimate the relative frequencies for each meaning, and then used these frequency estimates to define our entropy estimate:

$$S_n = \sum S_{n,m}$$

$$P_{n,m} = \frac{S_{n,m}}{S_n}$$

$$U_n = -\sum P_{n,m} \log_2 P_{n,m}$$

$$= \log_2 S_n - \sum P_{n,m} \log_2 S_{n,m}$$

where all sums are over the $M_n$ meanings, $m$. The resulting $U$ measures ranged from 0.31 (hold) to 2.24 (mole). Of these words, 220 were also included in the data collected by Twilley et al. (1994), which derived a similar measure of entropy from subject-reported associations, which we designate $U^A$. $U^A$ and $U$ were found to correlate significantly, $r(218) = .160, p = .018$. Some sample values for 2-meaning words are shown in Figure 1. Statistics for all the measures are included in Table 1.

[Table 1 about here.]

**Task**

The participants were outfitted with five head position indicator (HPI) coils, and their head shape, fiducial locations, and HPI coil locations were digitized before the lexical decision task. Participants lay in a dimly lit magnetically shielded room while the visual words were presented in pseudorandom order. The task was a continuous lexical decision task, in which the word or non-word stimulus was presented, prompting the subject to make a decision about lexicality and to press a button with the left index finger to respond ‘yes’ or with the left middle finer to respond ‘no.’ Stimuli were presented in lowercase, non-proportional Courier font, and subtended an average of 1.28° visual angle horizontally and vertically per character. Behavioral response times and accuracy data on
the lexical decision task were collected for each participant. During the task, neuromagnetic fields were recorded continuously by a 157-channel axial gradiometer whole-head MEG system (Kanazawa Institute of Technology, Kanazawa, Japan) at a 1000Hz sampling rate. As part of a separate session associated with a different experiment, a high-resolution T1-weighted anatomical MRI (MPRAGE sequence, $1 \times 1 \times 1$ mm) was obtained for each subject using a 3T Siemens Allegra head-only scanner. All acquisition was performed at the Center for Brain Imaging at New York University.

Analysis

The MEG data preprocessing steps were the same as those described in Solomyak and Marantz (2009b, 2009a) and Lewis, Solomyak, and Marantz (in press). First, raw data from 156 sensors were noise-reduced using data from three reference sensors located away from the subjects’ heads and the Continuously Adjusted Least-Squares Method (CALM; Adachi, Shimogawara, Higuchi, Haruta, & Ochiai, 2001) in Meg160 software (Yokogawa Electric Corporation and Eagle Technology Corporation, Tokyo, Japan). The noise-reduced MEG, head shape digitization, and sensor location data were imported into MNE (MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging, Charleston, MA) for additional processing. Subjects’ structural MRIs were reconstructed with the FreeSurfer software (CorTechs Labs Inc., LaJolla, CA). The data were then processed in MNE for estimation of each subject’s cortically constrained minimum-norm solution. Next, a source space consisting of 5124 activity points was created on each subject’s reconstructed cortical surface. At each source, activity was computed for the forward solution with the BEM (boundary-element model) method, which provides an estimate of each MEG sensor’s magnetic field. The forward solution and the grand average activity across all trials and subjects were used to compute the inverse solution, which estimates the most probable distribution of averaged MEG data across space and time. The data were next converted into noise-normalized dSPM (dynamic statistical parameter map) values (see Dale et al., 2000). The brain of each subject was morphed to a standard FreeSurfer brain, and activity from all trials and subjects were averaged and projected
on the standard brain. Regions of interest (ROIs) were then functionally defined on the standard inflated cortical surface, based on peaks in the grand average left hemisphere activation. ROI vertices were stored in label files, and then morphed back to each subject’s brain. For each subject, an inverse solution was computed for that subject’s ROI across all trials.

The M130 ROI was identified based on a peak in average negative activity (current directed inward on the cortex) in the posterior occipital region, consistent with Type I activity. The M170 ROI was based on a peak in average positive activity (directed outward) in the left-hemisphere occipito-temporal fusiform gyrus region between 150ms and 210ms post-stimulus onset. Lastly, the M350 ROI was based on grand average negative activation peaking around 300ms to 370ms in the left-hemisphere superior temporal and Sylvian Fissure regions. An extreme value count for each trial was made of data points with values falling two standard deviations from the overall mean. Trials were excluded from the analysis if their extreme value count exceeded the overall extreme value mean by more than three standard deviations.

For each ROI, the (directionally signed) activation from the MNE solution was averaged over the entire region to produce a time-course for each trial. In order to identify a representative time-course and peak response regions, the resulting time-courses were averaged over all included trials, and then over all subjects, thus timing was defined in absolute terms rather than relative to individual peaks. The time period around each peak during which the average response was within 5% of the total response range of the peak was selected for analysis. Within each included trial, the activation was averaged across this time period to get a single response amplitude per trial per region.

Standard group regressional analyses were used for behavioral and neural data, treating the first-level beta values as random variables, and testing whether the average response to each variable was significantly non-zero. The first-level design matrices consisted of the following regressors: a normalizing constant, word length \( L \), bigram frequency \( B \), form frequency \( F \), total meanings \( M \), total senses \( S \), and entropy \( U \). The regressors were orthogonalized in this order, thus \( L \) was
only mean-corrected, and $U$ had the projection of all the other regressors removed. All trials with non-words, incorrect (‘no’) responses, or response times over 2 seconds were excluded.

## Results

### Behavioral

Overall response accuracy was 92.9%. One subject had an accuracy rate of 77%, with the remaining subjects all performing over 90%. Fewer than 2% of trials were excluded for excessive response times over 2 seconds. In total, 5063 trials were included for analysis. Of these, 2298 had $U^A$ values available.

Mean response time was 711 ms with a standard deviation of 233 ms. There was a significant inhibitory effect of $U$ on response time, $t(12) = 2.91$, $p = .013$. There was also a significant effect of $U^A$ on response time, $t(12) = 3.48$, $p = .0046$. Under a fixed-effects model, the effect of $U$ was marginal, $r(5061) = .026$, $p = .06$, and $U^A$ was significant, $r(2296) = .064$, $p = .0023$. There were also significant facilitatory effects of: $F$, $t(12) = −6.27$; $M$, $t(12) = −3.53$; and $S$, $t(12) = −4.25$. By reversing the orthogonalization order of $M$ and $S$, we found that the total number of senses explained the significant modulation by meaning as well: $S$, $t(12) = −5.14$; $M$, $t(12) = −1.72$, n.s.

### Neural

After behavioral exclusions, an additional 73 trials were excluded from analysis based on MEG response characteristics. Based on average distributed source activity peaking at 133 ms in the left occipital cortex, the M130 ROI was defined over 4448 vertices, with a peak activity period of 124–143 ms. Identified responses in the occipito-temporal fusiform gyrus peaked at 211 ms, used to define the M170 ROI covering 6212 vertices and spanning 206–223 ms. Finally, the superior
temporal M350 ROI covering 4963 vertices was defined based on activity at the peak latency of 315 ms, the peak was found at 304 ms, and the period spanned 281–325 ms. The defined ROIs and average response time-courses are shown in Figure 2.

[Figure 2 about here.]

Within the M130 activation, only word length and bigram frequency were found to have significant effects on response amplitude, $t(12) = -2.84, p = .015$, and $t(12) = -2.99, p = .011$, while word form frequency approached significance. Within the M350 activation on the other hand, form frequency and entropy both had significant effects, $t(12) = -3.16, p = .0082$ and $t(12) = -2.34, p = .037$. All of these significant effects indicate increased magnitude (negative) activity with increased frequency and entropy. Using only the reduced data set for which $U^A$ was available this pattern of results held, except the effect of $U^A$ on M350 amplitude did not reach significance, $p = .085$. None of the M170 correlations was found to be significant, however in a post-hoc test, recognizing that our identified window was later than the M170 is commonly identified, we repeated the analysis at a single timepoint offset of 170ms and found a significant effect of length only, $t(12) = -5.37, p = .0002$. Within and between-subject permutation tests on both the whole design matrix and the entropy measures alone validated these significance levels (within .005). Repeating this analysis for each timepoint individually also reveals more details about the time-course in a more traditional correlational analysis (Figure 3).

[Figure 3 about here.]

Discussion

Many putative lexical measures that have been used to study the time-course of lexical resolution, most notably frequency, are purely derivable from syntactic properties of language, without re-
quiring any explicit semantic information. By parametrically varying a measure that is inherently
dependent on semantics and independent of other word-form measures we can help disambiguate se-
mantic and purely orthographic effects. The measure we used was based on the relative frequencies
of the distinct meanings of homographic homonyms, and assumes that these meanings are relevant
in some way to the lexical representation (e.g., as completely distinct lexical entries). Entropy is
an information-theoretic measure of uncertainty, or the amount of information (in bits) it requires
to disambiguate a particular meaning of a word from its polysemous visual representation. Since
lexical resolution necessarily implies resolving the meaning or meanings of a word, we expected
that this value would have correlates in the neural processes implementing this resolution.

The behavioral results confirmed that the entropy measure is relevant and inhibitory of word
recognition, with higher entropy resulting in longer response times. While we also confirmed the
general finding that ambiguity is advantageous in word recognition, with more senses resulting
in shorter response times, we only weakly replicated the behavioral results of Rodd et al. (2002)
in finding that homonymy had no significant effect after controlling for this effect of polysemy.
We also identified neural correlates of visual word recognition using MEG, and focused on three
distinct time and brain areas thought to participate in this process, associated with the M130, the
M170, and the M350 MEG responses. Looking for parametric effects of our various measures in
the amplitude of these responses, just as we did with response time, we found that, while the early
M130 response was sensitive to form measures, only the M350 activation showed effects of entropy.

These results are evidence that visual word processing involves some meaning-based resolution
beginning around 300 ms after presentation. This implies that earlier effects may be due solely
to orthographic properties that are derivable from the letter sequence itself. The null results for
earlier responses cannot prove the absence of semantic effects in the underlying neural activity or
that lexical resolution has not begun already, but do give credence to the idea that M350 activity is
the first to reflect semantics when taken with other findings of this nature (Solomyak & Marantz,
2009a; Lewis et al., in press).
While an appropriately controlled significant correlational finding is compelling regardless of the type of effect, the signs of the correlations are relevant to the network model. As predicted, we saw increased M350 activity to higher-entropy words, which implies that ambiguity introduces more inhibitory activity in the network, thus lengthening settling time and increasing net activity. This is consistent with the behavioral results. On the other hand, we also saw increasing activity to words with higher form-frequencies. This is contrary to the predictions of the simple network model, which stipulates that the baseline activity or prior probability for higher-frequency words is higher, thus requiring less evidence to activate and resulting in a faster resolution, as evidenced by the response time. However, it has also been found that M350 activity increases with morphological family frequency (Pylkkänen, Feintuch, Hopkins, & Marantz, 2004), which also makes sense within a network model framework, as it provides a measure of the number of prospective nodes that may receive activation by the target word. This is consistent with the findings of Lewis et al. (in press), which indirectly controlled for family size by using words with essentially no morphological families and found the traditional facilitatory effect of frequency, with reduced M350 activity for higher frequency bases. Indeed, when we considered the effect of cohort measures, we found an increase in M350 activity with increasing family size or frequency, but no significant effect of frequency after removing the variance explained by either of these variables, similar also to the findings of Solomyak and Marantz (2009b).

References


Figure Legends

Figure 1

Entropy function. Example entropy values for words with two meanings, based on relative frequency of less common meaning as proportion of total form frequency. Dictionary-based approximations calculated for this study are shown along with sample norms from Twilley et al. (1994).

Figure 2

(a) Selected regions of interest and (b) average response across subjects across included trials across each region. Indicated time ranges are analyzed peak regions having average response within 5% of peak.

Figure 3

Group correlation statistic timeseries from a regression of each individual timepoint separately. Also shown are the original average response (from Figure 2) and \( p = .01 \) (uncorrected) significance levels.
**Figure 1** *Entropy function.* Example entropy values for words with two meanings, based on relative frequency of less common meaning as proportion of total form frequency. Dictionary-based approximations calculated for this study are shown along with sample norms from Twilley et al. (1994).
Figure 2 (a) Selected regions of interest and (b) average response across subjects across included trials across each region. Indicated time ranges are analyzed peak regions having average response within 5% of peak.
Figure 3 Group correlation statistic timeseries from a regression of each individual timepoint separately. Also shown are the original average response (from Figure 2) and $p = .01$ (uncorrected) significance levels.
Table Legends

Table 1

Statistics of and correlations between the measures of interest across the 500 words (220 for $U^A$).
Table 1 Statistics of and correlations between the measures of interest across the 500 words (220 for $U^A$).

<table>
<thead>
<tr>
<th>statistics</th>
<th>$L$</th>
<th>$B$</th>
<th>$F$</th>
<th>$M$</th>
<th>$S$</th>
<th>$U$</th>
<th>$U^A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>4.222</td>
<td>2141</td>
<td>8.660</td>
<td>2.234</td>
<td>8.216</td>
<td>0.929</td>
<td>0.733</td>
</tr>
<tr>
<td>stddev</td>
<td>0.882</td>
<td>1101</td>
<td>1.478</td>
<td>0.536</td>
<td>4.694</td>
<td>0.312</td>
<td>0.467</td>
</tr>
<tr>
<td>min</td>
<td>3</td>
<td>75</td>
<td>6.3</td>
<td>2</td>
<td>2</td>
<td>0.310</td>
<td>0</td>
</tr>
<tr>
<td>max</td>
<td>6</td>
<td>6547</td>
<td>14.3</td>
<td>5</td>
<td>30</td>
<td>2.236</td>
<td>2.08</td>
</tr>
<tr>
<td>(bigram frequency) $B$</td>
<td>.537</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(log frequency) $F$</td>
<td>-.111</td>
<td>.032</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(meanings) $M$</td>
<td>-.063</td>
<td>-.004</td>
<td>.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(senses) $S$</td>
<td>-.044</td>
<td>-.011</td>
<td>.403</td>
<td>.227</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(est entropy) $U$</td>
<td>-.016</td>
<td>.040</td>
<td>-.128</td>
<td>.815</td>
<td>-.032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(norm entropy) $U^A$</td>
<td>.007</td>
<td>-.038</td>
<td>.036</td>
<td>.202</td>
<td>.087</td>
<td>.160</td>
<td></td>
</tr>
<tr>
<td>parts of speech</td>
<td>-.013</td>
<td>.003</td>
<td>.275</td>
<td>.094</td>
<td>.386</td>
<td>-.037</td>
<td>-.030</td>
</tr>
<tr>
<td>orth neigh size</td>
<td>-.767</td>
<td>-.202</td>
<td>.167</td>
<td>.105</td>
<td>.106</td>
<td>.045</td>
<td>.061</td>
</tr>
<tr>
<td>orth neigh freq</td>
<td>-.528</td>
<td>-.184</td>
<td>.132</td>
<td>.080</td>
<td>.143</td>
<td>.035</td>
<td>-.003</td>
</tr>
<tr>
<td>cohort size</td>
<td>-.153</td>
<td>-.007</td>
<td>.481</td>
<td>.029</td>
<td>.485</td>
<td>-.113</td>
<td>-.066</td>
</tr>
<tr>
<td>cohort freq</td>
<td>-.067</td>
<td>.050</td>
<td>.410</td>
<td>-.009</td>
<td>.204</td>
<td>-.038</td>
<td>.011</td>
</tr>
<tr>
<td>correlations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>